Amendments to the Specification:

Please amend paragraphs 27 and 30 as follows:

[00027] Accordingly, CMP systems 100 and 200 each consist of n processor cores 104a-104n and 204a-204n, respectively. For purposes of this disclosure, n may be any integer > 1, including 2, 4 and 8. For each embodiment 100, 200, the processor cores (104a-104n or 204a-204n, respectively) share a single shared memory system 150. It should be noted that, for each embodiment 100, 200 ~~200, 300~~, it is assumed that a latency mode is supported to allow a single non-speculative thread (the main thread) to persistently occupy one CMP core 104, 204 throughout its execution. Meanwhile, the remaining CMP cores 104x-104n are either idle or execute speculative helper threads to perform data prefetching for the main thread. In this manner, data fetching for a main thread may be accelerated by a speculative helper thread that executes concurrently on a different processor core 104 than the main thread.

[00030] Communications from helper threads on one core to a main thread running on another core may therefore suffer a higher latency ~~that~~ than communications among threads in an SMT environment, or other type of system where thread contexts share data in a lower cache close to the core. This increased latency may reduce the performance benefit that could otherwise be realized from helper threading.